DATA MANAGEMENT PLAN

Grant Agreement number	24DIT04
Project short name	ScanClouDT
Project full title	Traceability for industrial 3D digitalisation by advanced scanning systems
Data management plan	1st ☒ 2nd ☐
Confidentiality status	PU - Public, fully open

TABLE OF CONTENTS

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EURAMET. Neither the European Union nor the granting authority can be held responsible for them.

The project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

1 Data management plan

1.1 Data summary

Questions	Answers
1 Will you re-use any existing data and what will you re-use them for? State the reasons if re-use of any existing data has been considered but discarded.	The project has reviewed suitable artefacts suitable for the point cloud characterisation study in WP2. Some prominent artefacts have been characterised in previous research projects. These artefacts will be scanned again in the project, but reference data will be provided by the owners to help assessing the data quality of the projects results, for validating them, and for the providing point clouds of higher quality for the development of the quality metrics.. The participants have additional existing datasets on-hand that are not available to the entire consortium but can be used for validation and cross-checking to seek for explanations or confirmation in case of unexpected findings. So far, no datasets have been considered but discarded.
2 What types and formats of data will the project generate or re-use?	Original scanning raw data is stored in *E57* format. Most of the project data generated by physical and virtual laser scans (the digital metrological twins (D-MTs)) will be stored in *LAS/LAZ* (LIDAR Aerial Survey / LASzip) formats, the widely used binary format for an efficient storage of 3D point cloud data. The virtual artefact geometries will be generated based on CAD data: *.step, .stl, .obj, .bim, .xml* or ASCII. They will be processed using Blender (*.blend*). The Consortium will develop a bpy-based tool to generate a generic hybrid voxel format. This will be defined by the project and published in the gitlab repository. An exemplary processing chain will be set up for *.obj* and *.blend*. Data will be generated in the following formats: • Scan data: *.las/.laz* • Original scan data: *.e57* • Measurement and analysis results: ASCII *.csv*, *.las/.laz* • Graphics: *.jpeg, .pdf, .png* • Tables: *.xlsx, .csv* • Text: *.pdf, .txt, .docx* • Parameters, metadata: *.xml, .json, .sql* • Gitlab context (data publication e.g. docker, metadata): *.yaml, .xml, .sh* Similar data formats will be used as input (artefact geometries *.stl, .obj,* etc.) and output (point clouds *.las/laz, .e57,* etc.) to the simulation platform for reference data generation in WP3.
3 What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Experimental data will be used to develop and verify the D-MTs central to Objective 1 and to develop and verify the metric of Objective 2. Objective 3 literally calls for reference datasets. These will be generated and assessed in WP3. In the case studies, data will be generated and utilised to demonstrate the use of the outcomes of WP1 to WP3 for the measurement uncertainty assessment, and to derive Good Practice Guides. The latter two aspects are essential for the fulfilment of Objective 4. The generated and analysed data will support the development of new guidance and standardisation documents supporting Objective 5.
4 What is the expected size of the data that you intend to generate or re-use?	The expected data volume is >20 TB.
5 What is the origin/provenance of the data, either generated or re-used?	The existing data will originate from several sources, which will include: data from the scientific literature, real-world measurement data and data from simulations. The data collected from domestic properties will remain confidential and will not be included in the repository.
6 To whom might your data be useful ('data utility'), outside your project?	The project's data will be suitable for use by other research groups working on the following topics: surveying, geodesy, civil engineering, metrology, developers

	of AI methods in terrestrial laser scanning (TLS). It will be useful for quality control in surveying, aerospace, automotive industries, precision mechanical engineering, advanced manufacturing industry, instrument manufacturers of terrestrial laser scanners (TLS), TLS software developers Moreover, the dataset will be useful for standards committees including ISO/TC 172/SC6 *Optics and Photonics – geodetic and surveying instruments*, ISO/TC 213 WG10 *Coordinate measuring machines*, and ISO/TC 28/SC 2 *Measurement of petroleum and related products*.

1.2 Findable, Accessible, Interoperable and Re-usable (FAIR) Data

1.2.1 Making data findable, including provisions for metadata

Questions	Answers
7 Will data be identified by a persistent identifier?	Yes, if applicable, DOIs or commit/tag on Git repositories will be used.
8 Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.	Appropriate metadata will be used to describe text data used for journal publications, good practice guidelines and reports. The metadata will contain e.g. description keywords, date of deposit, author(s); the European Partnership on Metrology funding; the grant project name, acronym, and number. Where applicable, the metadata will include persistent identifiers (e.g. DOI) for related publications and research outputs.
9 Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?	Keywords will be provided in line with the FAIR principles.
10 Will metadata be offered in such a way that it can be harvested and indexed?	Zenodo complies with FAIR principles (https://about.zenodo.org/principles/). The metadata are indexed in a searchable resource. Metadata are licensed under CC0, except for e-mail addresses. All metadata are exported via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and can be harvested.

1.2.2 Making data accessible

Questions	Answers
Repository:	
11 Will the data be deposited in a trusted repository?	The data will be deposited in a trusted repository (Zenodo, Community ScanClouDT – 24DIT04).
12 Have you explored appropriate arrangements with the identified repository where your data will be deposited?	Zenodo is already working and correctly labels the datasets with a metadata scheme which is compatible with DataCite.
13 Does the repository ensure that the data are assigned an identifier? Will the repository resolve the identifier to a digital object?	Yes, there are unique identifiers for Digital object data (device zip files). Yes, the repository (Zenodo) can resolve the identifier to a digital object.
Data:	
14 Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in	All of the data that are needed to validate the results presented in scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. *Datasets which cannot be shared – voluntary restrictions* Other data may be made available on a case-by-case basis if it is relevant for third parties.

Questions	Answers
multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.	*The following data will not be made publicly available:* - Data obtained with the permission of third parties, but the third parties have not agreed to make the data publicly available. - Data that compromises the protection of a participant(s) intellectual property. - Pre-existing internal data of a participant, but the participant has not agreed to make the data publicly available. The level of data made available will also be considered, for example, pre-processed data will not be provided unless there is a clear reason for doing so.
15 If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.	The data used in scientific publications, posters and oral communications will be made available for re-use as soon as is reasonably possible. If necessary, open access will not be provided to some of the data/research outputs due to IPR considerations (e.g. whilst a patent application is pending). The data/research outputs will remain accessible for the lifetime of the repository.
16 Will the data be accessible through a free and standardised access protocol?	Yes, Zenodo provides well described conditions for free and standardised access (see http://about.zenodo.org/policies/).
17 If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?	There are no restrictions on the use of the published data, but users will be required to acknowledge the project and the source of the data in any resulting publications.
18 How will the identity of the person accessing the data be ascertained?	There is no need to ascertain the identity of persons accessing the data.
19 Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?	The ScanClouDT Consortium does not see the need to establish a Data Access Committee. If requests are deemed critical by partners and conflicts cannot be resolved by the parties involved, the Management Board will get involved. Ethical aspects and data security, including intellectual property requirements, will be considered as will access requests to personal/sensitive data. If necessary, some or all of a potential publication's data will be withheld. This will be decided in consultation with the relevant participant(s) (e.g. authors).
Metadata:	
20 Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?	In Zenodo, metadata will be made openly available. Metadata is licensed under CC0, except for email addresses. All metadata are exported by Zenodo via OAI-PMH and can be harvested.
21 How long will the data remain available and findable? Will metadata be guaranteed to remain available after data are no longer available?	The data will remain available and findable for the lifetime of the Zenodo repository, which is expected to be a minimum of 20 years. If data is withdrawn from Zenodo, the DOI and the URL of the original object are retained. In case of closure of the Zenodo repository, best efforts will be made by Zenodo to integrate all content into suitable alternative institutional and/or subject based repositories.
22 Will documentation or reference about any software be needed to access or read the data and will this be included? Will it be possible to	Most data will be in well-documented (open-source) formats readable with common software tools available to the public as per the answer to Question 2. Where necessary, reference to required software (versions) will be provided with the datasets. When original datasets require highly specialised, commercial software, the consortium will seek paths to provide information converted into non proprietary, well-documented additional formats like E57.

Questions	Answers
include the relevant software (e.g. in open source code)?	

1.2.3 Making data interoperable

Questions	Answers
23 What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?	The datasets will use the trusted repository's basic metadata schema for administrative data, which is compliant with the recommended standards used by DataCite (https://search.datacite.org/) and OpenAIRE (https://www.basesearch.net/). For individual datasets, the following discipline-specific vocabularies, standards, formats, and methodologies will be used: VIM, GUM, ISO 10360 series, ISO 9849, DCC schema and refType thesaurus.
24 In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow their re-use, refinement or extension?	Mapping will not be required as the terminology used will be chosen to be compatible with the existing literature.
25 Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?	Yes, the project's datasets that will be deposited in the chosen repository (e. g. Zenodo) might include qualified references to other datasets from the same project or from previous research.

1.2.4 Increase data re-use

Questions	Answers
26 How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?	A short README file (e.g. Markdown) will be provided together with the data, in order to enable data analysis and to facilitate data re-use.
27 Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement?	The data/research outputs will either be licensed under the latest available version of the CC BY license or a license with equivalent rights as set out in the Grant Agreement. Users will be required to acknowledge the consortium and the source of the data in any resulting publications.
28 Will the data produced in the project be useable by third parties, in particular after the end of the project?	Any data published in open-access journals will be usable by third parties after the datasets have been deposited in Zenodo. The data that do not relate to peer-reviewed publications will be made available for re-use on a case-by-case basis.

[1] A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)

Questions	Answers
29 Will the provenance of the data be thoroughly documented using the appropriate standards?	Yes, the provenance and context of the data will be thoroughly documented. Data will be accompanied by information on how they were captured, processed, analysed, and validated. Other information that enables interpretation and use will also be provided as well as licensing information.
30 Describe all relevant data quality assurance processes.	Data quality will be assured through several quality assurance procedures: - Repeated and comparison measurements. - Use of controlled vocabularies and standard terminology. - Metrological characterisation of the measurement set-ups. - Verification/validation of the data collected (if appropriate). - Provision of test results along with the data. - Peer-review of publications based on the data.
31 Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.	Allocation of resources The estimated costs for making the (data and) other research outputs FAIR are 2 500 € (personnel costs) (see question 34). The costs for making other research outputs FAIR are included in the project's budget and will be claimed if compliant with the Grant Agreement's conditions. The Project Management Board, in which all participants are represented, will have overall responsibility for managing other research outputs. Where feasible, long-term preservation will be ensured by depositing the other research outputs in repositories. The Project Management Board, with support from the participants, will decide on a case-by-case basis on which other research outputs will be deposited and for how long. *Security of other research outputs* All participants are either accredited to, or work in compliance with, the ISO 17025 standard on the "General requirements for the competence of testing and calibration laboratories". The participants will store other research outputs on their organisations' networks, which are protected by firewall, backups etc. Other research outputs will also be stored in the project's cloud drive (PTBbox) environment, with password-protected login. Deposition in public repositories will provide additional security as they have multiple replicas in a distributed file system which is backed up on a nightly basis. This project will not generate sensitive other research outputs. The other research outputs will be safely stored in open access repositories. *Ethical aspects* There are issues that could impact the sharing of other research outputs. - Information relating to other research outputs acquired from third parties, e.g. manufacturers, will not be shared without their explicit consent. - Information relating to other research outputs collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. The project will not share other research outputs with identifiable personal information. Sensitive information relating to the other research outputs will be collected, separated as soon as possible, and kept secure. Please also see the information provided in section 1.3 below.

1.3 Other research outputs

Questions	Answers
32 In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or	If not covered by peer-reviewed publications, the new parametrisation methods, and protocols produced by the project will be described and stored in the project's Zenodo repository.

Questions	Answers
re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).	
33 Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.	As far as possible, the FAIR data approaches specified in questions 7-30 above will be applied to the management of this project's other research outputs. This commitment will be met by releasing the protocols and parametrisation methods in a trusted repository.

1.4 Allocation of resources

Questions	Answers
34 What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?	The estimated costs for making the data and other research outputs Findable, Accessible, Interoperable and Re-usable (FAIR) are 2 500 € (personnel costs). These costs have been kept to a minimum by using a free repository (Zenodo) and by making only relevant data and other outputs FAIR.
35 How will these be covered? Note that costs related to research data/output management are eligible as part of the European partnership on metrology grant (if compliant with the Grant Agreement conditions).	The costs for making the data FAIR are included in the project's budget and will be claimed if compliant with the Grant Agreement's conditions. Long term preservation will be ensured by depositing the data within repositories (Zenodo). There are no costs associated with the long-term preservation of the data in these repositories.
36 Who will be responsible for data management in your project?	The participant(s) who generate and/or process the data will have overall responsibility for the management of data/research outputs and quality assurance. The coordinator will be responsible for coordinating updates to the data management plan. The participant(s) involved in the data generation/processing will be responsible for deciding on a case-by-case basis which data/research outputs will be kept and for how long. The participant(s) that produced the data will be responsible for organising backup and storage, archiving, and for depositing the data/research outputs within the chosen repositories.
37 How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?	Long term preservation will be ensured by depositing the data within repositories (Zenodo). There are no costs associated with the long-term preservation of the data in these repositories. The data will serve as information about methods and tools developed in the project. The data will increase in value over time due to its impact in a wide range of applications and because it will enable stakeholders to understand the methods developed in the project. Hence, the data will facilitate the uptake of the project's results. The coordinator together with the consortium will decide on a case-by-case basis what data will be kept and for how long.

1.5 Data security

Questions	Answers
38 What provisions are or will be in place for data security (including	*Data recovery and secure storage* All participants are aware on the need to store data redundantly, with

data recovery as well as secure storage/archiving and transfer of sensitive data)?	technologies to recover data. Each participant have generally implemented respective rules and practices in their institution. All NMIs are either accredited to, or work in compliance with, the ISO 17025 standard on the "General requirements for the competence of testing and calibration laboratories". TEKNIKER's dimensional labs are accredited according to the criteria of 17025, but TEKNIKER is also certified in accordance with UNE-EN ISO 9001. AKL-tec is accredited according to DIN EN ISO 9001 and currently in the accreditation process to the ISO/IEC 27001 'Information technology – Security techniques – Information security management systems – Requirements'. While not being formally accredited, the academic partners ENS Paris-Saclay, TUM, UCL, and WUT follow good scientific practice as required by national and international academic bodies like the German Science Foundation (DFG), e.g. Finally, Ommatidia as a young company with a data-based business model, has clear internal guidelines on and capabilities to maintain data safety and storage well-aligned with ISO 9001 rules and good academic practice. Primarily, the participants store data on their organisations' networks, which are protected by firewall, backup systems etc. Data of general interest and significance is also be stored in the project's collaborative environment (PTBbox), with password protected login. Deposition of published data in the Zenodo public repository will provide additional security as it has multiple replicas in a distributed file system which is backed up on a nightly basis. *Transfer of sensitive data* If needed, sensitive data will be transferred using the PTBbox (password protection, expiry times of links) or individual protected communication between participants.
39 Will the data be safely stored in trusted repositories for long term preservation and curation?	Yes, the data will be safely stored in the Zenodo open access repository. Zenodo and the underlying Invenio Framework for digital repositories were designed according to the Open Archival Information Systems (OAIS) reference model. Zenodo is working towards ISO 16363 certification.

1.6 Ethics

Questions	Answers
40 Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics report(s) and the ethics section in the Annex 1.	There are issues that could impact on data sharing. - Data acquired from third parties, e.g., manufacturers, will not be shared without their explicit consent. - Data collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. - The data from the stakeholder surveys will be made anonymous to comply with the General Data Protection Regulation (GDPR). No ethics issues are expected.
41 Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Informed consent for data sharing and long-term preservation will be included in the market and customer surveys, but the project has no plans to share data with identifiable personal information. If any sensitive data are collected they will be separated as soon as possible and kept secure.

1.7 Other issues

Questions	Answers
42 Do you, or will you, make use of other national / funder / sectorial / departmental procedures for data management? If yes, which ones (please list and briefly describe them)?	Data management will be compliant with: - The research data policy of the European Partnership on Metrology; - European laws about data security and the protection of privacy (e.g. GDPR); - Institutional guidelines;

	- Scientific community guidelines.

2 Open science: research data management

Statement	Put an X in the box to confirm	Or, list any exceptions to this
All participants have adhered to the requirements of the project's GA and CA with respect to open science: research data management (GA Article 17 and its Annex 5) for this reporting period	☒	